

Identification of Vigilance Lapses using EEG/EOG by Expert Human Raters

M. T. R. Peiris^{1,2,3}, R. D. Jones^{1,2,3,4}, P. R. Davidson^{1,3}, G. J. Carroll^{1,5}, T. L. Signal⁶, P. J. Parkin^{1,5},
M. van den Berg⁶, P. J. Bones^{1,2}

¹ Van der Veer Institute for Parkinson's and Brain Research, Christchurch, New Zealand

² Electrical and Computer Engineering, University of Canterbury, Christchurch, New Zealand

³ Medical Physics and Bioengineering, Christchurch Hospital, Christchurch, New Zealand

⁴ Medicine, Christchurch School of Medical and Health Sciences, University of Otago, Christchurch, New Zealand

⁵ Neurology, Christchurch Hospital, Christchurch, New Zealand

⁶ Sleep/Wake Research Centre, Massey University, Wellington, New Zealand

Abstract—It is critically important for certain occupational groups to remain highly alert throughout their working day. For safety reasons, it would be useful to automatically detect lapses in performance using EEG/EOG. Automating the detection process could be simplified considerably if we could mimic human experts. Surprisingly, it is unclear to what extent human EEG raters are able to detect lapses. Consequently, we undertook a study in which 4 expert EEG raters assessed the level of alertness of 10 air traffic controllers by observing a combination of their EEG and EOG while they performed a 10 min psychomotor vigilance task (PVT). They were specifically required to identify lapses or sleep episodes that might lead to a lapse in PVT performance. A reaction time ≥ 500 ms was defined as a PVT lapse. There was a total of 101 lapses (mean duration = 1.00 s). Of these, only 6 lapses were detected by one or more raters and all of these were marked as 'sleep'. Overall the human expert raters were unable to reliably identify lapses based only on EEG and EOG. This poor performance suggests an automated system would need to identify subtle features not overtly visible in the EEG.

Keywords—Alertness, drowsiness, lapses

I. INTRODUCTION

Tiredness and mental fatigue can lead to brief instances of falling asleep while engaged in some active tasks, like driving a motor vehicle [1, 2]. It is critically important for certain occupational groups, such as pilots, air traffic controllers (ATCs), and train drivers to remain highly alert throughout their working day. Diminished levels of alertness and lapses can have disastrous consequences including multiple fatalities. We would like to be able to detect these events automatically and in real time by analyzing EEG and EOG. As a first step we needed to determine the ability of human expert EEG readers to detect lapses. Consequently, we undertook a study in which 4 expert EEG raters classified the alertness level of 10 records (10 min each) chosen from 10 ATCs, using cues in the EEG and EOG.

We defined a lapse as a complete, temporary loss of responsiveness to a visual stimulus. A primary goal of this study was to determine whether experts could reliably identify lapses in visuomotor performance by observing cues in the EEG and/or EOG.

II. METHODOLOGY

A. Subjects

Neurophysiological and psychomotor performance data was recorded from 28 air traffic controllers (19M/9F; mean age 35.5 years, range = 26–56) at various times during their night shifts.

B. Apparatus

The subjects were connected to an ambulatory neurophysiological recorder (Embla by Flagstaff) and the following EEG channels were recorded: C4–A1, O2–A1, O2–Oz, O2–P4, and Oz–P4. Left and right EOG was recorded from the left and right outer canthus, with the left electrode positioned 1 cm up from the horizontal plane and the right 1 cm down and both referenced to the auricular reference (A1). This placement allowed the recording of both horizontal and vertical eye movements as well as slow eye movements. The data was digitized at 200 Hz using a 16 bit A-D converter.

We used a psychomotor vigilance task (PVT) [3] to measure the performance of subjects across the night shift. The PVT is a validated, reliable and sensitive test of vigilance and reaction time [4]. The test required individuals to respond as fast as possible to the presentation of digits on an LED display. The test was of 10 min duration and the inter-stimulus interval ranged between 2–10 s.

C. Procedure

Each ATC was involved in the study on four separate occasions. Two of these involved night shifts starting at 2230 and two at 2330. All subjects completed the PVT three times (mid-way through the shift, during a scheduled break, and at the end of the night shift) during each study

This research was financially supported by the Foundation for Research Science and Technology and the University of Canterbury, Christchurch, New Zealand.

night shift, giving a total of 12 PVT sessions per subject. The subject's EEG and EOG were recorded while they performed the PVT.

All reaction time (RT) responses ≥ 500 ms were defined as *lapses* [5]. All sessions were ranked according to the number of lapses, and the 10 sessions from independent subjects containing the highest number of lapses were selected to be rated by 4 experts.

Two teams of 2 raters were asked to classify the records using a continuous marker into the following categories: 1) Alert, 2) Light Drowsy, 3) Deep Drowsy, 4) Lapse in Consciousness (i.e., complete, temporary loss in responsiveness) other than sleep, and 5) Sleep. The teams defined their own rating criteria to classify the EEG and/or EOG.

We verified that the subjects were alert during the first 10 s of the PVT by inspecting the RTs of each session to ensure that they did not contain any lapses during the first 10 s. This provided the raters with a baseline alertness level for that subject. The EEG, EOG, and EEG+EOG of the subjects were presented to the raters in a random order. PVT data was not provided. Each expert rated the data independently of the other raters.

D. Analysis

None of the raters identified a category '4' event. Therefore this category was ignored for the remainder of the analysis. A rater scored a *hit* if the rating was marked as '5' (Sleep) for some period during a lapse. It was classified as a *false positive (FP)* if there were no lapses during the period the rater classified as '5', provided it did not occur during an inter-stimulus period. If this was the case, we could not definitively conclude that the rater made a false detection as there was no performance sampled during the inter-stimulus period.

The number of hits and FPs were calculated for EEG only, EOG only, and EEG and EOG rating data.

III. RESULTS

A. Lapse profile

TABLE I
LAPSE DETECTION BREAKDOWN FOR THE 4 RATERS (ASSESSING ALERTNESS VIEWING SIMULTANEOUS EEG AND EOG DATA)

Subject	No. lapses	Rater 1		Rater 2		Rater 3		Rater 4	
		Hits	FP	Hits	FP	Hits	FP	Hits	FP
1	31	0	0	0	0	2	1	0	0
2	21	0	0	0	0	2	0	4	0
3	17	0	0	0	0	0	0	0	0
4	12	0	0	0	0	0	0	0	0
5	5	0	0	0	0	0	0	0	0
6	4	0	2	0	0	0	0	0	0
7	4	0	0	0	0	0	0	0	0
8	3	0	0	0	0	0	0	0	0
9	2	0	0	0	0	0	0	0	0
10	2	0	0	0	0	0	0	0	0

There was a total of 101 lapses in the 10 subjects studied, with mean lapse duration of 1.00 s. The mean number of lapses across all subjects was 10.1 (range 2–31). The longest lapse was 11 s.

B. Lapse detection by human raters

Table 1 shows the number of lapses and the numbers of hits and false detections made by each rater using cues in the EEG and EOG. A total of 6 lapses were correctly detected by at least one rater, and there was agreement between two raters on just 2 of these lapses.

None of the raters had hits or FPs using the EOG only. However, for the case of EEG only, 7 lapses were correctly detected by one or more raters. There was agreement between 2 different rater pairs on 2 lapses and 3 raters agreed on a single lapse.

C. Relationship between lapses and EEG/EOG ratings

We calculated the overall proportion of lapses beginning within epochs marked at each EEG/EOG rating level. The assigned level at the start of most lapses was Alert (mean proportion across raters = 0.76), and far fewer started during epochs marked as Light Drowsy (14%), Deep Drowsy (8.4%) or Sleep (1.5%). Since Alert was the most common rating level, we also calculated average lapse rates within each rating level. We divided the total number of missed stimuli within a level by the total duration of that level. This analysis was applied across data from all sessions and was completed separately for each rater. The mean lapse rates across raters increased with rating level. Mean \pm SD lapse rate was 0.95 ± 0.23 lapses/min for Alert, 1.08 ± 0.63 lapses/min for Light Drowsy, 1.86 ± 2.15 lapses/min for Deep Drowsy and 4.62 ± 5.34 lapses/min for Sleep.

IV. DISCUSSION

Detection of vigilance lapses by human expert raters was poor, with only 6 of 101 lapses being detected, and with inter-rater agreement on only 2 of these. Slightly more lapses were detected viewing EEG alone and none were detected when viewing the EOG alone, suggesting EOG was of little value.

Lapses occurred more frequently during intervals rated as Sleep or Deep Drowsy but, since most of the data was rated Alert, the majority of lapses occurred during alert epochs.

Even though only those sessions containing the most lapses were selected, more than half the selected subjects had less than 5 lapses during the entire session. In most cases these lapses were less than 1 s. Our results suggest that there are no overtly visible cues in the EEG and EOG during these short lapse events which enable raters to detect them consistently.

There were a number of deficiencies in the data set which may have hindered the rating and the consequent

detection of lapses. We could not obtain an accurate estimate of lapse duration because PVT stimuli appeared at discrete intervals, so performance was not assessed during the inter-stimulus period. Consequently, it is possible that the actual lapse commenced several seconds in advance of the stimulus onset, but was only apparent after the occurrence of the stimulus and the consequent prolonged RT. For this reason we did not count ratings of 'sleep' occurring during the inter-stimulus interval as FPs.

Although sufficient for sleep rating, the EEG data only contained 5 channels. Some raters may only be familiar with full-head EEG and may have found it difficult to adjust to fewer EEG channels. For example, there were no frontal EEG channels recorded, so no eye movement cues were presented to the raters during the EEG only rating.

There was also no independent measure of the subject's level of alertness, such as video recording. This would have been of considerable value in determining whether a prolonged RT was due to a diversion of attention or a behavioural microsleep episode.

Despite these shortcomings, our results demonstrate that detecting lapses based on EEG and EOG is not a trivial task.

for expert EEG raters. This suggests that, unlike an automated sleep stager, an automated lapse detection system may need to identify features that are not visible in the EEG.

REFERENCES

- [1] F. Sagberg, "Road accidents caused by drivers falling asleep," *Accid. Anal. Prev.*, vol. 31, pp. 639-649, 1999.
- [2] P. Cummings, T. D. Koepsell, J. M. Moffat, and F. P. Rivara, "Drowsiness, counter-measures to drowsiness, and the risk of a motor vehicle crash," *Inj. Prev.*, vol. 7, pp. 194-9, 2001.
- [3] J. W. Powell, "PVT-192 and analysis software reference manual," Unit for Experimental Psychiatry, University of Pennsylvania, Philadelphia 1999.
- [4] D. F. Dinges, M. M. Mallis, G. Maislin, and J. W. Powell, "Evaluation of techniques for ocular measurement as an index of fatigue and as the basis for alertness management," National Highway Traffic Safety Administration, Washington, DC. Final Report DOT HS 808 762. 1998.
- [5] D. F. Dinges, F. Pack, K. Williams, K. A. Gillen, J. W. Powell, G. E. Ott, C. Aptowicz, and A. I. Pack, "Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night," *Sleep*, vol. 20, pp. 267-77, 1997.